ChinaXiv合作期刊 图书情报工作
第62卷 第24期 2018年12月

KRDS科学数据管理成本–收益模型调研与分析*

卫军朝　李思雪　刘盼攀
上海大学图书情报档案系　上海 200244

摘要：[目的/意义]对KRDS模型的理论和实践进行调研分析，为我国开展科学数据管理提供指导和借鉴。[方法/过程]通过文献调研、网络调研，对JISC资助的KRDS模型和基于该模型的国外高校科学数据管理实践进行总结，详细解析KRDS模型在科学数据管理成本分析中的应用，得出对我国开展科学数据管理的启示。[结果/结论]KRDS模型是一个通用的、系统的架构，基于KRDS对科学数据管理成本进行分析，能比较完整地涵盖和预测到各个管理环节。KRDS模型可为我国开展科学数据管理提供参考方向，规范管理流程，细化管理成本和收益。

关键词：科学数据管理　KRDS模型　作业成本法
分类号：G250
DOI：10.13266/j.issn.0252－3116.2018.24.013

以数据密集型计算为特征的科学研究第四范式正在兴起，科学数据成为支撑科研发现的重要资源，越来越多的机构和组织开始进行科学数据的管理工作，以有效支撑基于数据的知识发现。从科学数据管理的具体操作来看，科学数据管理是一系列围绕科学数据的活动总和，包括数据的组织、备份、存档、共享、出版和安全管理等，这些活动保障了科学数据的使用和再利用，是支撑基于数据的科学发现的重要基础[1]。对这些科学数据管理活动的成本以及可能产生的收益进行研究和分析，有助于我们更加有效地开展科学数据管理，保障科学数据管理的可持续性，为我国开展科学数据管理活动提供参考和借鉴。

国外已有机构开展科学数据管理成本的研究和实践工作。A. S. Palaiologk等利用作业成本法分析了荷兰数据存档与网络服务(Data Archiving and Networked Services(DANS))的科学数据管理和长期保存成本[2]，该方法侧重于将影响科学数据管理成本的因素划分为人力成本和非人力成本，再藉由作业活动分解成本驱动因素。欧洲社会科学数据档案联盟(Consortium of European Social Science Data Archives(CESSDA))开发了社会科学数据管理和长期保存的成本－收益分析工具包(Cost-Benefit Advocacy Toolkit)[3]，它侧重于社会科学数据管理，目的是帮助机构识别、理解、评估数字资源管理成本和收益。此外还有一些已经比较成熟的数字资源长期保存成本评估模型，例如荷兰国家档案馆2005年提出了T-CMDP模型[4]，主要针对电子表格、邮件等类型的数据，成本核算时间跨度仅涉及现在和未来；美国国家航空航天局在2008年改进了其NASA-CET模型[5]，侧重于空间和多维数据的保存与管理；以及伦敦大学和英国图书馆2010年提出的LIFE3模型[6]，它尤其适用于图书、报纸等电子资源管理成本对象，且部分涉及成本核算的固定值来自对图书资源的实证调研，缺乏适用性等。其中，由英国联合信息系统委员会(Joint Information Systems Committee(JISC))资助的KRDS(Keeping Research Data Safe)模型[7]，支持多种学科类型的科学数据管理成本和收益核算，成本核算时间跨度长(过去、现在与未来)，成本驱动因素划分相对合理全面，同时还弥补了收益分析方面的缺失，故本文选取KRDS模型作为研究对象。

国内的研究主要集中于数字资源长期保存成本与收益的研究。例如苏小波[8]、肖英[9]等对数字资源保存的成本驱动因素进行了分析；杨鹤林[10]、臧国全[11]等对图书馆数字资源保存中的成本与效益关系进行分析；孙超[12]、肖秋会[13]、李海涛[14]等对各种国外已经

* 本文系上海市软科学重点项目“科学数据管理成本－效益分析及可持续运行模式研究”(项目编号:18692104300)研究成果之一。
作者简介：卫军朝(ORCID：0000－0002－3113－1455)，讲师，博士，E－mail：weijch@ shu.edu.cn；李思雪(ORCID：0000－0002－8668－5439)，硕士研究生；刘盼攀(ORCID：0000－0002－8810－2161)，硕士研究生。
收稿日期:2018－06－20　修回日期:2018－08－09　本文起止页码:93－103　本文责任编辑:杜杏叶

成熟的数字资源保存成本模型进行比较了分析与评价。其研究对象是广义的数字资源,包括科技文献、科学数据、音频、视频等。专门以科学数据为研究对象,通过对科学数据管理活动的分析来进行科学数据管理成本－效益的研究尚未见到。

数据密集型科研范式下,科学数据管理成为大多数科研机构的常规活动,对科学数据管理的成本效益进行分析,能够帮助我们识别科学数据管理中的关键活动,促进机构开展高效的科学数据管理。本文以此为切入点,通过比较国外已有的科学数据管理成本收益实践与模型(前文已述),选取 JISC 资助的专门应对科学数据管理成本－收益分析的 KRDS 评价模型为调研和分析对象。该模型具有支持多类型科学数据管理成本－收益分析,成本驱动因素全面,时间跨度长,适用性强,能够为国内的相关应用提供清晰的思路等特点。通过文献调研、网络调研的方法,通过对 KRDS 项目的产出报告,包括 Keeping Research Data Safe (KRDS1)[15]、Keeping Research Data Safe 2 (KRDS2)[16]、I2S2/KRDS Benefits Analysis Tools Project[17] 以及 Keeping Research Data Safe 2 (KRDS2) Project website 网站[18] 的调研,以及 KRDS 模型在各高校的应用报告,例如剑桥大学[19]、伦敦国王学院[15]、南安普顿大学[20] 等高校应用 KRDS 模型的项目报告,结合相关研究论文,从构成要素、成本驱动因素、成本核算框架、收益分析框架、应用实践等多维度深度解析 KRDS 模型。希望能够洞悉 KRDS 科学数据管理成本－收益模型,解析构成科学数据管理成本的主要活动,总结影响科学数据管理成本的驱动因素,归纳该模型的应用程序,为我国开展科学数据管理成本和收益分析提供参考。

1 KRDS 模型概述

1.1 KRDS 模型发展历程

KRDS 模型是由 JISC 资助的关于评估科学数据管理和保存成本与收益的项目成果,主要用于帮助机构识别和确定科学数据管理与保存的成本,并使其意识到管理和保存活动会带来的收益。该项目通过调查和研究高等教育机构(Higher Education Institutions (HEIs))在科学数据管理和保存方面的中长期成本与收益情况,开发了相关的理论、工具、方法,就有关问题对英国高等教育基金会(Higher Education Funding Council for England(HEFCE))和机构进行了指导,保证了科学数据管理的可持续性发展。

KRDS 模型基于生命周期成本法来分析科学数据管理的成本活动,通过确定生命周期中科学数据管理成本活动、成本驱动因素(影响活动成本的变量,如数量、格式等)和资源(员工时间、设备等),构建综合的成本核算框架,帮助管理机构理解、识别、分析管理活动中的成本,同时它还提醒管理者进行收益分析和建立相关价值链。该模型经历了三个发展时期:第一期(KRDS1)完成于 2008 年,该阶段建立了成本分析模型的主体,探讨了构成科学数据管理成本的主要因素;第二期(KRDS2)完成于 2009 年,对模型进行了修改和优化,识别并分析了与长期数据保存有关的收益因素;第三期主要是对成本收益模型与工具进行推广和使用,将研究成果转向实践领域[7]。

1.2 KRDS 模型的构成

KRDS 模型主要由三部分构成:作业模型,成本驱动因素以及成本核算框架(见图 1)。作业模型用于识别具有成本影响的科学数据管理活动,并将它们排列到一个分为作业、子作业的层次结构中;KRDS 成本驱动因素,即关键变量(例如工资水平或通货膨胀率),它们影响保存活动的成本。成本驱动因素分为两大类:经济调整和服务调整;KRDS 成本核算框架将成本(员工或设备等)和作业持续时间(1 年、2 年等)联结起来,形成一种接近于透明成本核算法(Transparent Approach to Costing(TRAC))的综合成本核算模型。TRAC 广泛应用于英国的 165 所高等教育机构,用来核算高校的教学、科研以及其他主要活动的经费[21]。

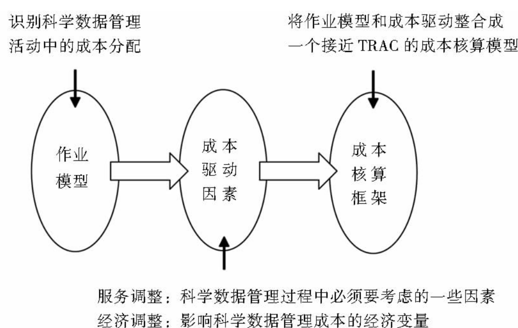


图 1 KRDS 模型构成[22]

KRDS 作业模型有助于识别消耗资源的活动;成本驱动因素能够帮助科学数据管理机构识别作业活动中影响成本的经济变量,以及管理和服务过程中必须要考虑的一些因素,例如数据格式等,这些变量最终需要数据管理机构与数据提交方协商确定;成本核算框架将作业模型和成本驱动因素联结在一起,方便机构进行学习、参照和成本核算。由此,数据保管机构便可根据 KRDS 模型,结合本

卫军朝,李思雪,刘盼攀.KRDS科学数据管理成本-收益模型调研与分析[J].图书情报工作,2018,62(24):93-103.

地情况,识别消耗资源的管理活动,确定成本驱动因素,核算和分析整个科学数据管理过程中的成本。

2 KRDS 作业模型和成本驱动因素

2.1 作业模型

科学数据管理是科研人员所从事的对科学数据的管理活动,这些活动消耗资源,产生成本,这些成本构成了科学数据管理的成本。因此,识别科学数据管理的行为和活动,是厘清科学数据管理成本的首要步骤。KRDS 作业模型就是用于识别科学数据管理活动的工具,它由一套层次结构组成,包括预归档、归档两个活动阶段,以及附加的支持服务和物业管理。预存档阶段涉及所有与数据创建和管理相关的活动,这些活动是在存档之前。存档阶段包括一系列将科研数据归档到由高校或其他机构运行的档案存储库中的活动。这两个阶段都与科学数据的生命周期成本有关。支持服务包含的活动可以支持预存档或存档阶段的活动,它们通常是关于财务、IT 和其他公共服务的基础设施。而物业管理则包含对建筑物和其他基础设施的管理。

KRDS 作业模型基于对 LIFE、NASA CET、OAIS 和 TRAC 模型的研究与分析,经过不断地修改和扩展,从 KRDS1 演化到 KRDS2。KRDS 作业模型的细分列表如图 2 所示[23]:

预存档阶段	宣传	开展	创建	存档阶段	获取	处理	摄取	归档存储	长期保存规划	先发创新	数据管理	访问	支持服务	管理	公共服务	物业管理
		项目设计;数据管理计划	协商知识产权/许可/伦理		选择(数据)	向存档机构转移;销毁	接收提交数据;质量检查	接收摄取的数据;管理存储层级;更换介质	跟踪指定的用户社区;跟踪技术的发展	制定团体数据标准和最佳实践;共享保存系统和工具	管理和维护数据库	检索和命令;生成信息包发给用户;交付响应		日常行政/综合管理;行政/管理支持	操作系统服务;网络服务;网络安全服务;软件许可和硬件维护	
		基金申请;项目实施/启动	生成科学数据;生成描述性元数据;生成用户文档		协商提交协议;存储帮助/提交数据向导		生成存档信息包;生成管理元数据;生成/更新描述性元数据和用户文档	灾难恢复;误差校验;提供访问副本	制定保存战略和标准;制定信息包策划和迁移计划;制定和监控外包保存的服务级别协议	与供应商合作	执行查询;生成报告;数据库更新	用户支持;新产品生成		制定政策和标准	物理安全;公共事业(水电);物资库存和物流;相关人员培训	
			开发定制软件;数据管理;生成数据存档包				统筹/协调更新;参考链接		进行长期保存;生成保存元数据							

图 2 KRDS 作业模型

KRDS 作业模型将科学数据保存管理活动分为两个阶段:预存档阶段和存档阶段,以及配套的支持服务和物业管理。预存档阶段主要负责创建科学数据及将其转换为数据存档,其中要考虑到如数据格式、元数据等影响数据生成和获取的因素,为此就要指定相关的数据保存和共享计划,生成关于数据的描述性信息、用户文档,并与数据创建者协商所使用的格式和逻辑结构等,此外管理机构还会为数据提交者和创建者提供存档训练和支持。

存档阶段是科学数据长期管理的主要阶段,初始阶段要制定相关的数据选择政策,与数据创建者协商提交协议;之后会将数据转移到知识库或保管机构,并安全销毁没有被选中进行长期管理的数据;数据管理机构会提供适当的存储容量和设备来接收数据,并将数据格式转换为存档所需的格式;在数据保存归档的过程中,会生成相关的管理元数据、描述性元数据、用户文档等,并且会提供可以更新归档文件内容的机制以及原始数据文本解释的语义链接;在数据归档管理时,要维护数据完整性,及时增改和删除信息;当用户进行访问和查询请求时,能够实时为用户生成结果集、报告并提供相关培训和帮助。在数据创建、管理和共享等过程中,要保持与数据消费者、生产者的交互,以跟踪需求和技术的变化,随时更新保管技术和策略,新的保存技术、工具、标准的开发,需要各个机构共同参与开发和制定。

支持服务需要行政人员提供日常的支持与控制,建立和维持数据保管标准与政策;此外还需要提供数据平台的软件接口、分布式应用程序、不间断的水电资源、日常办公消耗品和人员培训与发展等。物业管理和服务费用则涉及房屋租赁、空间管理和维修,在

KRDS 模型中，根据功能的不同（如实验室/非实验室），将其归为与其他公共服务分离的成本因素，并按可变利率核算费用。

这些科学数据管理活动，消耗人力资源、资本资产，存档过程中相关的数据量、数据格式等因素会影响资源的使用情况，从而影响成本。管理机构可以借助该作业模型识别管理活动，为进一步分析成本做准备。

2.2 成本驱动因素

为了进行成本核算，保管机构在明确管理活动后，需要确定成本的驱动因素，这些因素以及其调整和变化会通过资源与资产的消耗而导致成本的产生和变动。成本驱动因素包括多种不同的变量，这些变量能影响科学数据管理成本。KRDS 模型中将这些变量分为两类，即经济调整类和服务调整类。KRDS 模型列出这些成本驱动因素[15]，帮助科学数据管理机构在进行成本分析时，识别管理活动中会影响成本的变量，通过具体变量的调整和变化来核算资源的消耗、成本和收益。

（1）一般要素。在进行各活动阶段具体成本变量的识别之前，KRDS 首先确定了那些会对管理成本产生整体性影响的因素，它把这些因素统称为一般要素。这将有助于识别后续具体活动阶段的成本变量。一般要素在表 1 中列出[15]：

表 1 一般要素

一般要素	说明
数据集等级和保存目标	高等教育机构的多数数据集仅供项目团队使用，有时也只用于极少数外部用户。保存目标是保证科研数据能持续安全存储，并且有足够的描述性信息，以便数据恢复。
控制未来成本	通过限制一些服务调整类因素，机构可以控制未来成本的复杂性和不可预测性。
（行动）时机	在预存档阶段中生成描述性元数据和用户文档，而不是在存档的摄取活动中生成新的/升级不足的元数据和文档，对于节省成本尤其重要。
成本依赖、联系和“涟漪效应”	上面的行动时机是成本依赖的一个例子，任何保存科学数据成本模型都要考虑到它的存在。
对工作负载和进程时间调度的敏感性	人力资源不易快速地适应总存储量的变化，或工作负荷的短期波动，特别是保管机构对数据资料何时到达或处理速度几乎没有控制能力时。
保存技术的发展和商用现货的可获得性（COTS）或成熟的开源软件/和社区标准和最佳实践（“先发创新”）	技术的发展和商用现货的可获得性或成熟的开源软件应用在不同的生命周期阶段中都将对成本产生重大影响。通常，这些开发适合于外部资助和协作来完成。在作业模型中，这被作为先发创新阶段的一部分。

（2）经济调整和服务调整。KRDS 模型将一系列具体成本驱动因素分为经济调整和服务调整两类。经济调整类的成本驱动因素主要是指影响科学数据管理成本的经济变量，当它们发生变化时，一系列资源和资产的成本核算就会随之变动。例如，通货膨胀率、投资回报率等。管理机构要与数据提交方商定这些变量，以便之后的成本核算。

服务调整类的成本驱动因素则是指在科学数据管理、保存过程中，各个作业阶段在提供服务或要求时，必须要考虑的一些因素，例如数据量、存储格式、用户数量等等。科学数据管理机构需要在进行存档活动之前制定相关标准和预期，使得数据提交方能够按标准规章和流程提交数据，以便核算成本并确保后续的保存服务。表 2 列出了经济调整和服务调整两类成本驱动因素[15]：

表 2 影响科学数据管理成本的变量

分类	作业阶段	变量
经济调整		通货膨胀/通货紧缩；折旧；融资和投资的回报
服务调整	通用类	人力成本和劳动率；作业持续时间；启动和作业操作阶段；自动化的水平
	获取、处理和摄取	用户数量；储存的数量、方式和频率；文件格式的数量、复杂性和类型；数据量；元数据、文档、伦理和知识产权；处理、验证和校准的级别；取消访问成本
	存档存储、保存规划、数据管理	保留期；管理和更新；版本和副本的数量；存储介质（容量，成本）；存档介质监测
	访问	用户和用户社区的数量；标准或自定义接口；用户支持的水平；访问控制；访问的数量和容量/体积；访问/分配方法；服务响应时间；加工过的产品

（1）经济调整。经济调整类因素需要管理机构、数据提交方以及资助机构进行协商确定。通货膨胀率通常适用于诸如员工等成本的核算；通货紧缩率通常适用于一些设备类的成本核算；折旧的计算通常是依据时间的推移或资产的活跃（或使用）水平；而融资和投资的回报则包括融资成本和最低限度的保留盈余。

它们的变化将会影响一系列资源、资产的成本核算，从而确定经济类变量变得尤其重要。

(2)服务调整。在所有作业活动中，尤其要注意人力成本，它应包括工资、保险和养老金(退休金)，该成本是科学数据管理活动中的主要成本。在案例研究中，管理活动的 70% 甚至更多的成本与人力成本有关，而从历史上看，这些费用一直被视为管理成本的主要组成部分。也因为员工成本的重要性，自动化水平对总成本来说就是一个重要变量，它的影响程度将取决于其可实现的规模经济。当然，为了能够获得和模拟通货膨胀/通货紧缩的费用和调整成本，还需要记录作业的持续时间。除此之外，在不同的活动阶段成本的侧重类型也会不同，例如在启动阶段强调安装系统基础设施的固定成本，而在后续作业阶段则强调随时间推移运行容量的可变成本。

在获取、处理和摄取阶段要特别注意存档文件格式的专有性将会对成本产生很大影响，非专有格式可以简化获取和迁移程序，降低管理风险和成本。对于许多新领域和应用程序，只有专有格式可用，在这种情况下，一个至关重要的因素就是支持数据导出(或导入)为这些专有格式。此外，还要注意，描述性元数据、伦理和知识产权许可的作业时机是非常重要的，在预存档阶段进行这些活动会大大降低成本，如果这些活动要在存档阶段进行调整，不仅会提高费用，数据的价值可能也会显著下降；值得注意的是，科学数据管理的大部分成本在获取和摄取阶段而不是在长期存档和保存阶段。

在存档存储、保存规划、数据管理作业中，数据的保留期越长，就需要更多的保护措施，以确保数据的完整性和访问性，因此会造成更高的总成本。此外，服务水平也会影响成本，例如吞吐量、误码率、硬件替换频率等。同样，介质迁移频率、样本采集频率等都会涉及人员和设备的利用与消耗，产生成本。在用户访问过程中，成本是潜在地且变动性最大的阶段。根据访问级别和方法的不同，成本是弹性的，例如通过网络访问和由工作人员直接处理的请求，产生的成本将非常不同。此外，为了满足用户高速访问和生成特定数据产品等需求，都会产生费用。

基于 KRDS 的作业模型，科学数据管理机构能将其管理活动清晰化、标准化，利用成本驱动因素列表，机构便能在其存档前识别影响成本的变量，明确或预测资源和资产的消耗，完成管理成本的预算，并便于资助机构、社群进行投资和收益分析。

3 KRDS 成本核算框架

成本核算框架是一个核算科学数据管理成本的简化而通用的框架，它将 KRDS 作业模型中的要素和成本驱动因素联结在一起，为科学数据管理机构开展成本核算提供了便利。成本核算框架涵盖了预存档、归档和支持服务以及基于 TRAC 的成本类别，此外该框架中还增加了 KRDS 模型中需要明确的存档费用和外包成本。通常，作业模型有助于识别管理活动中所需的资源或资产；经济调整类成本驱动因素有助于明确经济变量或运行情况带来的成本变动，在开展成本核算时必须要考虑该类因素；服务调整类因素则有助于识别和调整活动中具体涉及的资源变量。根据该框架(见表 3)，机构能够清晰地明确各个作业活动中涉及成本的因素和变量，进而顺利开展成本核算。

表 3　成本核算框架[15]

预存档	存档	支持服务	持续时间(1 年、2 年等)
人力成本	人力成本	人力成本	
设备成本	设备成本	设备成本	
差旅费	差旅费	差旅费	
消耗品	消耗品	消耗品	
物业成本	物业成本	物业成本	
间接成本	间接成本	外包	
外包/存档费用	外包		

成本核算框架中的成本类别基于 TRAC 的成本分类方法，涉及员工、设备、旅行、消耗品、物业成本和间接成本。在完整的 TRAC 分类法中，员工成本又被分为直接产生或直接分配成本。要注意的是，KRDS 模型规定，在进行员工和其他成本的核算时，需要数据管理机构和资助机构在协商后，将经济调整类成本驱动因素纳入核算的考量中。成本核算框架只是一个简化模型，在具体实践中，要根据本地情况进行更详实的划分。

4 KRDS 收益分析框架

对管理科学数据的成本进行分析核算，不足以评估特定科学数据管理活动的经济可行性。成本分析应该伴随收益分析，换句话说，就是在维持科学数据长期存在和可获得性时，还要预期从投资中能产生多少价值。许多与科学数据管理有关的研究经济问题的文献都集中在成本/收益的成本方面，对管理活动为利益相关者所带来的收益的关注相对较少。通常，科学数据管理的投资会带来收益被认为是常识，或者用过于宽泛的术语来表达，从而限制了关于收益的分析和研究。

对未来科学数据管理的经济可行性进行分析，需要成本与收益相权衡。不幸的是，衡量收益常常是相当具有挑战性的，尤其是当这些收益不容易量化时。作为收益分析的第一步，制定一些重要的维度来阐明科学数据管理可能产生的收益非常重要。KRDS 模型中提出了这样一种收益分析框架[16]，它来自于对英国数据档案、牛津大学等科学数据管理机构的案例研究，从三个不同的维度分析科学数据管理可能会产生的收益，见表 4。虽然对许多收益开展量化很困难，但利用 KRDS 收益分析框架，至少能够让数据管理机构对管理科学数据所带来的收益有更加深入的了解。

表 4　KRDS 收益分析框架

维度 1	
直接收益	间接收益（避免成本）
新的研究机会	没有数据的再创建
学术交流/访问数据	没有丢失未来的研究机会
重新定位和再利用数据	更低的未来保存成本
提高研究效率	为新用户重新定位数据
激励新网络/合作	重新定位方法
知识向产业转移	新用户的使用
技术基础/技能组成	保护早期投资的回报
提高生产力/经济增长	
科研/科研诚信的验证	
实现授权	
维度 2	
短期收益	长期收益
对当前研究人员和学生的价值	
没有数据在短期流通中丢失	对未来研究人员和学生的价值；时间越长，随着收集量和临界质量的增加，价值也会增加
精选数据的短期再利用	
为数据密集型研究提供安全存储	
期刊论文数据的可用性	
维度 3	
私人收益	公共收益
对研究/存档发起人/资助者的收益	为未来研究投入
对研究人员的收益	激励新的研究
履行义务	促进新公司和高技能就业/职业
增长可见性/引用	
使研究商业化	

直接/间接、短期/长期和私人/公众三个收益分类维度，旨在帮助机构思考与长期保存科学数据有关的收益，以便更好地评估它们对于管理成本的相对权重。显然，还有其他维度可以添加到这个框架中，需要进行更多的实践和研究来丰富每个维度中的收益。虽然这些收益中的许多是很难量化的，在某些情况下甚至是不可能的，然而，即使只是通过定性的方式来表达收益，也有助于提高资助者和其他决策者的意识。至少，这种收益框架将有望激励管理机构、社群和社会对长期保存和管理科学数据的收益有更深入的了解，并且有助于阐明成本/收益等式的收益方面。

5　KRDS 模型的应用

目前国外许多机构和高校都开展了科学数据管理服务，尤其是英国高校，很多都基于 KRDS 模型进行了科学数据管理的成本和收益分析，如剑桥大学、伦敦国王学院、南安普顿大学等。本文总结了这些机构利用 KRDS 模型开展科学数据管理成本分析的步骤，见表 5。

从表中可以看出，高校在应用科学数据管理作业成本模型时，大致都分为以下几步：在确定具体的作业范围、文本提交格式和提交协议等后，首先明确管理过程中那些会对成本产生整体性影响的因素；之后结合自身情况应用和改进 KRDS 作业模型，按照 TRAC 方法划分成本或直接按照资源将成本划分为人力成本和资产成本；最后分析和预测每个作业活动的具体成本。根据各个高校科学数据管理成本 – 收益的研究情况、代表性以及相关数据的可获得性，选取伦敦国王学院、剑桥大学和南安普顿大学三个典型的案例，对 KRDS 科学数据管理成本 – 收益模型的应用进行详细分析。

5.1　伦敦国王学院 KRDS 成本模型应用

该案例研究基于 11 年来摄取和管理复杂科学数据集和 e-Research Center 为伦敦国王学院（King's College London（KCL））建立科学数据管理和保存的经验。KCL 科学数据管理被包含于一个更大的项目，即整合虚拟研究环境（Virtual Research Environment（VRE）），以支持 e-Research 的实践和运作。具体成本分析步骤如下：

（1）分析一般管理问题。在开展科学数据管理成本核算之前，KCL 根据 KRDS 成本驱动因素的一般要素列表，首先明确管理过程中那些会对成本产生整体性影响的因素。如首先要确定 3 个关键的工作范围：KCL 科研数据的内容和质量；旧系统与 VRE 的整合；VRE 的用户需求。此外，还需要考虑该项工作带来的其他问题：除了支持自身的科研需求之外，学院应当对 KCL 的科学数据负多大责任；数据是否能开放获取；对更大的科研社区负什么责任；如何满足与开放存取相关的额外费用；跨学科研究越来越趋向于国际合作，因此，对跨机构合作项目产生的科学数据如何合理分配责任；如何满足 KCL 科学数据管理和保存费用，以及 TRAC 是否能提供可行的解决方案等。

表 5　KRDS 模型应用案例

成本分析研究案例/机构/高校	数据管理对象	分析步骤
伦敦国王学院	高校数据存储库;国家数据中心;艺术与人文学	1. 利用 KRDS 成本驱动因素的一般要素列表,分析科学数据管理中对成本有整体性影响的问题; 2. 依据 KRDS 成本核算框架和 TRAC 要素,分配各个作业阶段的成本; 3. 开展具体成本分析与核算。
剑桥大学	高校数据存储库;化学,社会人类学	1. 利用 KRDS 成本驱动因素的一般要素列表,分析科学数据管理中对成本有整体性影响的问题;2. 依据 KRDS 作业模型,识别各作业活动涉及的具体人员,基于此分析人力成本; 3. 开展具体成本分析与核算。
南安普顿大学	国家数据中心;高校数据存储库;化学,海洋学	同上
牛津大学	高校数据存储库	同上
考古数据服务中心	国家数据中心;考古学	同上
英国数据档案	国家数据中心;社会和历史学	同上
收益分析研究案例/南安普顿大学,国家结晶学数据服务	国家数据中心;化学	1. 将该机构进行科学数据管理的活动划分为三个时期; 2. 进行原始数据和迁移数据的成本分析; 3. 利用 KRDS 收益分析框架进行收益分析。
英国数据档案	国家数据中心;社会和历史学	同上

(2)应用 KRDS 成本核算框架。在协商确定一般管理问题后,KCL 科学数据管理机构依据 KRDS 成本核算框架,利用 TRAC 要素进行了成本分配。首先将 KRDS 作业模型各阶段中所涉及的影响成本的要素划分为:直接产生、直接分配和间接三个类目。该机构在开展成本核算时提供审计跟踪,以指示直接花费在数据管理上的费用;直接分配成本基于管理设备的全部经济成本(Full Economic Cost(FEC))[15]。按照 TRAC 的指导方针,所有的成本都是直接分配的,而不是间接的。具体呈现见表 6。

(3)成本分析。KCL 科学数据管理机构依据自身情况,成本分析主要划分为人力成本、硬件成本分析。关键员工(成本)为归档管理人员(协助活动,薪资£ 45 000),半职系统管理员(安装和管理软硬件,全职薪资£ 24 000),集合专员(制定和实施管理、保存科学数据的工作流程与标准,薪资£ 35 000)。硬件成本是在 2005 年购买的东西,包含 15TB 的存储空间、一个磁带库和一个允许终端用户访问的分布式服务器[15]。此外,在具体的实践活动中还要考虑到设备存储容量、组织 IT 环境、资助者的需求变化等带来的影响,包括数据集和存档的大小、复杂性和类型等的改变,并且要在整个实践过程中要保证对设备进行持续维护和及时升级。

(4)预测数据管理成本。预测基于成本“峰值”的概念,它是指随着存档的扩大,需要更多的设备和存储容量来管理不断增加的数据量,因此成本也随时间而增加。类似的,随着存档的扩大,每年的收集数量增加,从而需要额外的工作人员。人力成本是根据收集人员每年处理 30 个集合的摄取量和保存情况而定的,其中包含 10 个“简单”的集合(标准格式和元数据的图像或文本)和 20 个复杂的集合(图像、视频以及相互链接的文档等),其中,员工 20% 的时间还会用于一般的任务[15],例如审查和更新许可证协议、标准等。

5.2　剑桥大学 KRDS 成本模型应用

剑桥大学案例研究的主要对象是 DSpace@ Cambridge 和 Department of Chemistry's Unilever Centre for Molecular Science Informatics。DSpace@ Cambridge 存储库最初是一个合作项目,从 2003 年到 2006 年由剑桥－麻省理工学院资助,主要目标是建立 DSpace 软件平台,使它作为剑桥大学机构存储库[19]。它接受各种格式的数字内容,主要是包含一些传统科研出版物的图片和数据。现含有化学信息学、考古学和人类学田野调查等学科的数据。其成本核算的主要步骤为:

(1)分析一般管理问题。同样,在开展科学数据管理成本核算之前,剑桥大学的科学数据管理机构根据 KRDS 成本驱动因素的一般要素列表,明确了管理过程中那些会产生整体性影响的因素。在确定剑桥数据管理的中长期成本时,首先需要考虑一些重要问题:选择和/或评估流程,创建充足的元数据(需要大量的人力成本),支持的格式/版本,不同格式的保存计划,真实性和可用性需求,以及可持续性等。此外,还要考虑到要雇佣数字保存领域的专家来协调整个保存活动的费用。

(2)应用 KRDS 作业模型。剑桥大学科学数据管理机构依据 KRDS 作业模型,识别各作业活动涉及的具体人员,在此基础上开展的成本分析主要集中于人力成本分析。详细描述见表 7。

表 6 伦敦国王学院 KRDS 成本核算框架

作业模型	TRAC	作业模型	TRAC	作业模型	TRAC	作业模型	TRAC	作业模型	TRAC
预存档阶段开展(项目设计;数据管理计划;基金申请;项目实施/启动)	直接产生	协商提交协议 推广和存储帮助	直接产生 直接产生(与项目直接相关) 直接分配(一般的推广和支持)	归档存储 接收摄取的数据;提供访问副本;管理存储层级;更换介质;灾难恢复;误差校验	直接产生 直接分配	先发创新(制定团体数据标准和最佳实践;共享保存系统和工具;与供应商合作)	直接分配	支持服务管理(日常行政/综合管理;客户账户;行政/管理支持)	直接分配
创建(协商知识产权/许可/伦理;生成科学数据;生成描述性元数据;生成用户文档;开发定制软件;数据管理;生成数据存档包)	直接产生	处理(向存档机构转移;销毁)	直接分配	长期保存规划(跟踪指定的用户社区;跟踪技术的发展;制定保存战略和标准;制定信息包策划和迁移计划;制定和监控外包保存的服务级别协议;进行长期保存)	直接分配	数据管理(管理和维护数据库;执行查询;生成报告;数据库更新;管理和维护数据库;执行查询;生成报告;数据库更新)	直接分配	公共服务(操作系统服务;网络服务;网络安全服务;软件许可和硬件维护;物理安全;后勤;公共事业(水电);物资库存)	直接分配
存档阶段 获取 选择(数据)	直接分配	摄取(接收提交数据;质量检查;生成存档信息包;生成管理元数据;生成描述性元数据和用户文档等)	直接产生	生成保存元数据	直接产生	访问 检索和命令;交付响应;用户支持;新产品生成 生成信息包发给用户	直接分配 直接产生	物业管理 耗材 差旅和必须品	直接分配 直接分配 直接产生

表 7 剑桥大学 KRDS 成本核算框架

阶段	人员	时间
启动阶段,保存项目最初的文件和数据,需要相关人员和数据提交方就存储方法、格式等方面进行协商。	项目人员与存储库管理员和系统管理员协商	每个项目需要 2 名 8 级全职员工花费 2 至 5 天 取决于各种因素,例如格式和项目工作人员对于 DSpace 是否熟悉
创建阶段,创建描述性元数据和用户文档是资源密集型的任务,研究者利用指导方针和专业支持进行操作,元数据越标准化效率会越高。	高校研究服务部法律组归于间接成本 联络人员(6 级)负责元数据映射	周期性 一次性工作,可能需要重新访问 项目开始的 1 –5 天 8 级和 6 级岗
获取阶段,需要大量的人力资源,但 DSpace 把选择内容的工作交由提交方来完成;扩展支持是资源密集型的,在生命周期的早期确定获取的数据、相关文档和软件是很重要的。	高校法律小组就提交条款意见 存储库经理(8 级)和支持及联络人员(6 级)	标准条款(一次性的)+额外的关于具体内容和使用的谈判(周期性),需要 5 天 >20% 的持续工作
转移阶段,数据存档在高校和外部管理机构之间转移,需要管理机构之间就技术和法律层面进行协商并统一标准。	存储库经理和系统管理员制定转移计划以及每年维护计划	一次性转移
摄取阶段,是成本最密集的阶段,但也是最有可能通过自动化来降低成本的阶段,取决于采用的摄取方法和是否采用自动化。	DSpace 系统管理人员(8 级)摄取数据 DSpace 系统经理和开发人员集成图像内容管理系统 作为 JISC 资助项目的一部分进行集成开发	周期性 10% 的持续工作 2 名 8 级岗工作 3 个月 3 年以上的£ 300000 项目资助
存档存储阶段,技术发展是影响该成本的关键因素,异地存储也会花费额外费用。		
保存规划,该阶段要制定数字资料生命周期管理规划,定期评估其长期价值、格式和风险等,确保其长期可用。	数字化及数字保存专家(8 级)与 DSpace 小组共同负责	周期性 全职 8 级岗
先发创新	存储库经理 DSpace 开发人员	周期性 2 名 8 级岗

相关数据来源:剑桥大学科研数据管理平台 http://www.data.cam.ac.uk/repository

(3)成本分析。剑桥大学数据管理机构在进行成本核算时,所涵盖的成本数据[19]主要分为以下几个方面:

- DSpace@Cambridge 的工作人员。DSpace@Cambridge 团队(3 名 8 级全职岗员工;1 名 6 级全职岗员工);数字化和数字保存专家(8 级全职岗)。但并没有指出具体的薪资核算方法。

- 硬件投资，DSpace@Cambridge 投资了大约 150TB 的镜像存储，成本约为£ 176,293.82。

5.3 南安普顿大学成本-收益模型应用

这个关于科学数据管理收益分析的案例研究，是基于南安普顿大学化学系所拥有的纵向成本信息。案例包含了 KRDS 模型中预存档和存档阶段的几个主要作业：开展项目设计和数据管理计划；创建描述性元数据，数据管理和生成数据存档包；数据选择和存储支持；向其他机构转移和销毁数据以及接收数据提交包等。具体步骤如下：

(1)划分时期。该案例是关于保存(原始存储格式)和迁移(新存储格式)成本的比较研究，数据来自国家结晶学服务，时间跨度为 1970-2009 年，在此期间，实验仪器、计算能力和数据存储介质(例如：纸，数字视频光盘等)已经彻底改变。考虑到这些变化的因素，该机构以技术之间进行转换为依据，例如：个人计算机的引入，新一代的仪器，或者在线存储的出现，将管理活动大致分为三个时期(1970-1990,1990-2000 和 2000-至今)[20]。

(2)成本分析。南安普顿大学数据管理机构的成本分析主要集中在保存成本和迁移成本方面。国家结晶学服务的实验结果是一个晶体结构，它是收集原始实验数据(Raw Data)的产物，并将其加工成结果数据(Result Data)。需要注意，实验室使用当前设备生成晶体结构的成本是£ 328，然而在 1970 年到 1990 年间重新创建一个结构的成本大约是六十倍。最明显的是，管理原始数据的成本约为总数据(原始+结果)管理成本的 70%[20]。因此，对原始数据的管理，而不是结果数据，是晶体结构数据管理的重要环节。

(3)收益分析。该案例研究强调的收益与 KRDS 收益分析框架相一致，见表 8。之后便根据表中得出的具体收益，在 KRDS 作业模型上详细叙述每个活动相对应的收益，确定收益类型、利益相关者以及收益实现年限，设定收益影响权重，从而形成一套完整的价值链。

6 对我国开展科学数据管理成本-收益分析的启示

第四范式正在兴起，为保障以数据为驱动的研究能顺利进行，对科学数据的管理已成为关键的研究和实践领域，在该管理活动中首要解决的是成本问题。国外高校在进行科学数据管理时，都会先确定管理阶段和作业，据此分析管理成本，为进一步管理活动做投资预算与方案准备。

表 8 南安普顿大学 KRDS 收益分析框架

维度 1	
直接收益	间接收益(避免成本)
因为实施正确和有用的元数据保存，提高科研生产力； 将知识转移到后继学者上； 对知识体系有更大贡献； 知识转移导致商业化的发现增加； 履行出资人授权的工作	早期投资的保护
维度 2	
短期收益	长期收益
对研究者： 在分析过程中恢复原始数据的能力； 在传播的早期阶段为原始数据提供来源链的能力； 提高了研究成果的可见度	保存禁运或未公布的数据(目前估计约占研究产出的 80%)[20]； 在收集并使其公开之后，恢复精确数据要便宜很多； 用下一代软件重新解释数据的能力
维度 3	
私人收益	公共收益
能够为未来管理个人研究数据，以便今后可以利用	增加资金价值； 增加知识转移； 增长可用于挖掘和新科学的数据体系

国内一些高校也渐渐开始重视科学数据管理，复旦大学于 2011 年成立了人文社会科学数据中心，通过这一跨部门、多学科联合机构的设立，该校人文社会科学数据资源能够实现集中建设、整体规划、统一管理和充分共享[24]，2013 年复旦大学与哈佛大学 Dataverse Network 系统签署了全面合作协议[25]，进行引进和二次开发，搭建了复旦大学社会科学数据平台，可以免费注册并上传、下载相关数据集；2014 年，北京大学图书馆开始对“研究数据管理平台”进行前期调研和平台选型，最后同样基于哈佛大学 Dataverse 系统搭建了开放研究数据平台[26]，并可供校内外用户进行数据集的免费上传和下载、共享等服务。国内高校虽然都在搭建数据中心，但几乎都是基于国外成熟的系统进行二次开发，现多为免费服务，且没有明确和详细的数据提交流程和成本核算模型与规定，关于科学数据管理成本-收益方面的研究和实践依然不够全面和完善。从以上模型研究和案例分析来看，我国在开展科学数据管理成本-收益分析时，应注意以下几个事项：

6.1 明确科学数据管理流程

根据国内数据平台目前的现状，在进行科学数据管理时，应该首先明确科学数据管理流程，提前制定数据管理规划；规定数据提交格式和提交协议，明确管理过程中那些会对成本产生整体性影响的因素等；确定数据收集和开放等级，保存目标；并且要特别关注工作负载量和进程时间，确保员工能够及时适应存储量的变化和工作负荷的波动。经研究发现，通过多个机构

合作建立科学数据管理中心,联合开发软件、追踪技术发展、制订数据管理计划、共享协议等,将会极大降低管理成本,提高数据共享效率。复旦大学、北京大学在建立其科学数据中心时,都调研和联合了多个部分和研究机构,但也仅仅是在本校范围,最多辐射当地的部分其他高校,规模效应还不够大;在其平台运行过程中,没有明确的数据提交协议和管理流程,这将会对保存成本和收益分析造成很大的困难。

6.2 应用成熟的成本分析模型

目前,国内的科学数据中心并没有制订出规范的成本分析模型,也没有很好的借鉴和改进国外成熟的成本－收益分析模型;各个科学数据管理机构在开展成本分析时问题不断,缺乏清晰的作业阶段划分、活动识别,成本驱动因素识别不当等。因此,科学数据管理机构或数据中心应结合自身情况,构建、改进和应用国外成熟的成本分析模型或,如 KRDS 成本分析模型,识别出适合本机构的作业活动,包括预存档和存档阶段的各种管理活动,并确定各个作业中影响成本的驱动因素,涉及数据格式、数据量、用户量等变量。其后,将作业活动与驱动因素连结起来,形成规范的本地化的成本分析模型,改善国内科学数据中心的成本分析现状。

6.3 协作开展成本核算

国内高校已搭建的科学数据管理平台现均处于免费服务阶段,用户在提交、下载与共享数据时,没有明晰的收费协议,平台内部也没有公布相关的成本核算步骤和方法,财务核算和平台服务相互分离,容易造成部分成本的缺失或重复核算。平台的科学数据管理成本核算应由相关技术人员和财务人员协作开展,该阶段首先要明确各个作业活动的成本类别。经案例研究发现,一般国外的科学数据管理机构按照两种方式来划分成本,一种根据 TRAC 方法,将活动阶段的成本划分为直接产生或直接分配;另一种是直接按照消耗的资源类型划分,通常分为人力成本和资产资本,这种方式需要管理机构确定各个活动阶段所需的员工类型和级别。要注意,在开展成本核算时,要考虑经济调整类的变量对资源和资产产生的影响,这将在很大程度影响整体成本。

6.4 注重科学数据管理人才的培养及实现自动化

从科学数据管理费用来看,人员支出占到了科学数据管理费用的大部分,在一些领域,甚至达到 90%。可见,科学数据管理人才是数据管理的关键因素。国外非常重视此类人才的培育,已成功实施了许多培育项目,而我国就图书情报专业来说,还未曾有与科学数据管理相关的课程,科学数据管理人才的培养已经迫在眉睫。此外,从数据保存作业的角度来看,数据摄取及其之前管理活动的人力成本占 55% 左右。然而实现高度自动化的操作,可以大大降低该部分人力成本。科学数据管理机构应当实时跟踪技术的发展,及时更新软硬件,以保证高效的数据摄取和管理工作,降低管理成本。

参考文献:

[1] Managing and sharing data [EB/OL]. [2018-04-13]. http://www.data-archive.ac.uk/media/2894/managingsharing.pdf.

[2] PALAIOLOGK A S, ECONOMIDES A A, TJALSMA H D, et al. An activity-based costing model for long-term preservation and dissemination of digital research data: the case of DANS [J]. International journal on digital libraries, 2012, 12(4): 195-214.

[3] CONSORTIUM OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES. CESSDA SaW [EB/OL]. [2018-04-13]. https://www.cessda.eu/Projects/All-projects/CESSDA-SaW.

[4] KEJSER U B, NIELSEN A B, THIRIFAYS A. Cost model for digital preservation: cost of digital migration [J]. The international journal of digital curation, 2011, 6(1): 225-267.

[5] NASA. CET V2p4 tech description 080915 [EB/OL]. [2018-04-14]. http://opensource.gsfc.nasa.gov/projects/CET/index.php.

[6] LIFE. An introduction to the third phase of the LIFE project [EB/OL]. [2018-04-14]. http://www.life.ac.uk/3/docs/life3_report.pdf.

[7] BEAGRIE C. Keeping research data safe: cost-benefit studies, tools, and methodologies focusing on long-lived data [EB/OL]. [2018-04-15]. https://www.beagrie.com/krds.php.

[8] 苏小波, 常娥. 数字资源长期保存的成本影响因素分析[J]. 图书与情报, 2011(1): 20-24, 44.

[9] 肖英. 数字信息保存成本研究[J]. 图书馆杂志, 2004(11): 14-17.

[10] 杨鹤林. 数据监护:美国高校图书馆的新探索[J]. 大学图书馆学报, 2011(2): 18-21, 41.

[11] 臧国全, 李思思. 数字保存项目的投资决策研究——基于项目投资成本不确定性的分析[J]. 图书馆理论与实践, 2014(8): 36-41.

[12] 孙超, 吴振新. 国外数字资源长期保存成熟度模型及其分析与评价[J]. 图书情报工作, 2017, 61(1): 32-39.

[13] 肖秋会, 许晓彤, 卞君妍. 数字保存成本模型 LIFE 和 CDL-TCP 比较与评价[J]. 图书情报工作, 2017, 61(18): 12-18.

[14] 李海涛. 会计成本视角下的公共图书馆数字保存成本模型[J]. 图书馆论坛, 2017, 37(2): 75-82.

[15] Keeping research data safe1 [EB/OL]. [2018-04-23]. https://www.webarchive.org.uk/wayback/archive/20140615221657/http://www.jisc.ac.uk/media/documents/publications/keepingresearch-

datasafe0408. pdf.
[16] Keeping research data safe2 [EB/OL]. [2018－04－28]. https://www. webarchive. org. uk/wayback/archive/20140615221405/http://www. jisc. ac. uk/media/documents/publications/reports/2010/keepingresearchdatasafe2. pdf.
[17] I2S2 idealised scientific research activity lifecycle model [EB/OL]. [2018－04－30]. https://www. beagrie. com/krds-i2s2/.
[18] Keeping research data safe2 -a JISC -funded project [EB/OL]. [2018－05－02]. https://www. beagrie. com/jisc/.
[19] UNIVERSITY OF CAMBRIDGE. Research data management [EB/OL]. [2018－05－01]. http://www. data. cam. ac. uk/repository.
[20] Southampton data survey: our experience and lessons learned [EB/OL]. [2018－05－01]. http://www. disc - uk. org/docs/SouthamptonDAF. pdf.
[21] HIGHER EDUCATION FUNDING COUNCIL FOR ENGLAND. Transparent approach to costing an overview of TRAC [EB/OL]. [2018－04－15]. http://www. jcpsg. ac. uk/guidance/downloads/Overview. pdf.
[22] KRDS user guide [EB/OL]. [2018－04－17]. https://www. beagrie. com/static/resource/KeepingResearchDataSafe_UserGuide_v2. pdf.
[23] Detailed version of KRDS2 activity model [EB/OL]. [2018－04－20]. https://www. beagrie. com/KRDS2_Activity_Model_detailed. doc.
[24] 刘霞，饶艳. 高校图书馆科学数据管理与服务初探——武汉大学图书馆案例分析[J]. 图书情报工作，2013，57(6)：33－38.
[25] 复旦大学社会科学数据平台 [EB/OL]. [2018－05－05]. https://dvn. fudan. edu. cn/home/index. jsp.
[26] 北京大学开放研究数据平台 [EB/OL]. [2018－05－05]. http://opendata. pku. edu. cn/.
[27] UNIVERSITY OF ESSEX. UK data archive [EB/OL]. [2018－04－14]. http://www. data-archive. ac. uk/.
[28] KRDS benefits framework worksheet [EB/OL]. [2018－05－02]. https://www. beagrie. com/KRDS_BenefitsFramework_Worksheetv1word_July2011. doc.
[29] Benefits impact worksheet [EB/OL]. [2018－05－03]. https://www. beagrie. com/Benefits% 20Impact% 20Worksheet_v1excel_July2011. xls.
[30] Value chain and benefits impact worksheet [EB/OL]. [2018－05－03]. https://www. beagrie. com/KRDS-ValueChainImpactTool_% 20Worksheet_v1excel_July2011. xls.
[31] 卫军朝，蔚海燕. “数据资产框架(DAF)”视角下的机构数据资产审计调研与分析[J]. 图书情报工作，2016，60(8)：59－67，92.
[32] WHEATLEY P，AYRIS P，DAVIES R，et al. LIFE：costing the digital preservation lifecycle [J]. 现代图书情报技术，2008(1)：69－74.
[33] SMITH M K，BARTON M，BASS M，et al. DSpace：an open source dynamic digital repository [J] D-lib magazine，2003，9(1)：10－17.

作者贡献说明：

卫军朝：负责总体架构和统稿；
李思雪：负责调研和论文撰写；
刘盼攀：内容修正。

Investigation and Analysis on KRDS Model of Cost and Benefits for Research Data Management

Wei Junchao　Li Sixue　Liu Panpan
Department of Library, Information and Archives, Shanghai University, Shanghai 200244

Abstract: [**Purpose/significance**] With the fourth paradigm of data-intensive scientific discovery emerging, research data management has been an important prerequisite for data-intensive scientific discovery. By the survey of Keeping Research Data Safe model, this paper aims to provide a reference for the research data management in China. [**Method/process**] This paper analyzed the JISC-funded Keeping Research Data Safe (KRDS) and the KRDS-based practice of overseas universities by literature research and Web-based survey, then analyzed the characteristics, steps and experiences and lessons learned of research data management cost analysis based on KRDS. Finally, it proposed specific measures for research data management cost analysis in China. [**Result/conclusion**] KRDS model is a universal and systematical framework. It can fully involve and predict the key cost of each activity. KRDS could provide a direction, standardize the management process, and refine management costs and benefits in China.

Keywords: research data management　KRDS model　activity based costing